Data Analyst Intern at Data Glacier
Week-9 : Deliverables

Project: Cross-Selling_Recommandation

Name: Kumkum Chakraborty

University: Dr.B.R.Ambedkar University

Email: kumkumchakraborty2016@gmail.com

Country: U.S.A

Specialization: Data Analyst

Internship Batch: LISUM41

Date: 02/03/2025

Table of Contents:

1.Problem Description:

XYZ credit union in Latin America is performing very well in selling the Banking products (e.g.: Credit card, deposit account, retirement account, safe deposit box etc.) but their existing customer is not not buying more than 1 product which means bank is not performing good in cross selling (Bank is not able to sell their other offerings to existing customer). XYZ Credit Union decided to approach ABC analytics to solve their problem.

2.Business Understanding:

The bank aims to increase cross-selling by analyzing customer demographics and financial behaviors without using machine learning. By Understanding income levels, age distribution, and product usage, the bank can offer tailored financial products like mortgages,investments,and pensions to the right customer. This will help improve customer engagement and product adoption. The project involves data inspection,cleaning,exporatory analysis, recommendations for data driven decision-making

Project Cycle

WEEK	DATE	PLAN
Week-7	02/19/2025	**Business Understanding**
Week-8	02/26/2025	**Data Understanding**
Week-9	03/02/2025	**Exploratory data analysis**
Week-10	03/09/2025	**Feature Engineering and model Building**
Week-11	03/16/2025	**Model Evaluation**
Week-12	03/23/2025	**Presentation**
Week-13	03/30/2025	**Document the Challenges**

Data Understanding

```
[5]:  print(train_data.head())

     fecha_dato  ncodpers ind_empleado pais_residencia sexo  age  fecha_alta  \
0  2015-01-28   1375586            N              ES    H   35  2015-01-12
1  2015-01-28   1050611            N              ES    V   23  2012-08-10
2  2015-01-28   1050612            N              ES    V   23  2012-08-10
3  2015-01-28   1050613            N              ES    H   22  2012-08-10
4  2015-01-28   1050614            N              ES    V   23  2012-08-10

   ind_nuevo antiguedad indrel  ... ind_hip_fin_ult1 ind_plan_fin_ult1  \
0          0          6      1  ...                0                 0
1          0         35      1  ...                0                 0
2          0         35      1  ...                0                 0
3          0         35      1  ...                0                 0
4          0         35      1  ...                0                 0

   ind_pres_fin_ult1 ind_reca_fin_ult1 ind_tjcr_fin_ult1 ind_valo_fin_ult1  \
0                  0                 0                 0                 0
1                  0                 0                 0                 0
2                  0                 0                 0                 0
3                  0                 0                 0                 0
4                  0                 0                 0                 0

   ind_viv_fin_ult1 ind_nomina_ult1 ind_nom_pens_ult1 ind_recibo_ult1
0                 0               0                 0               0
1                 0               0                 0               0
2                 0               0                 0               0
3                 0               0                 0               0
4                 0               0                 0               0
```

Data Shape

```
In [6]:  train_data.shape
Out[6]:  (13647309, 48)
```

Data Info

```
In [11]:  train_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13647309 entries, 0 to 13647308
Data columns (total 45 columns):
 #   Column                        Dtype
---  ------                        -----
 0   record_date                   object
 1   customer_id                   object
 2   employee_status               object
 3   country_of_residence          object
 4   gender                        object
 5   age                           object
 6   customer_since                object
 7   new_customer_index            object
 8   seniority_months              object
 9   primary_relationship_type     object
 10  last_primary_relationship     object
 11  customer_type_last_month      object
 12  residence_flag                object
```

DATA INFO 2

```
 19  active_customer_flag        object
 20  household_income            object
 21  customer_segment            object
 22  savings_account             object
 23  current_account             object
 24  derivada_account            object
 25  payroll_account             object
 26  junior_account              object
 27  mas_particular_account      object
 28  particular_account          object
 29  particular_plus_account     object
 30  short_term_deposit          object
 31  medium_term_deposit         object
 32  long_term_deposit           object
 33  e-account                   object
 34  funds                       object
 35  mortgage                    object
 36  pensions                    object
 37  loans                       object
 38  tax_payments                object
 39  credit_card                 object
 40  securities                  object
 41  home_account                object
 42  payroll                     object
 43  pension                     object
 44  direct_debit                object
dtypes: object(45)
memory usage: 4.6+ GB
```

Changing Column Name

All the column name has been changed for better understanding

In [9]:

```python
column_mapping = {
    'fecha_dato': 'record_date',
    'ncodpers': 'customer_id',
    'ind_empleado': 'employee_status',
    'pais_residencia': 'country_of_residence',
    'sexo': 'gender',
    'age': 'age',
    'fecha_alta': 'customer_since',
    'ind_nuevo': 'new_customer_index',
    'antiguedad': 'seniority_months',
    'indrel': 'primary_relationship_type',
    'indrel_1mes': 'last_primary_relationship',
    'tiprel_1mes': 'customer_type_last_month',
    'indresi': 'residence_flag',
    'indext': 'foreigner_flag',
    'canal_entrada': 'customer_acquisition_channel',
    'indfall': 'deceased_flag',
    'tipodom': 'address_type',
    'cod_prov': 'province_code',
    'nomprov': 'province_name',
    'ind_actividad_cliente': 'active_customer_flag',
    'renta': 'household_income',
    'segmento': 'customer_segment',
    'ind_ahor_fin_ult1': 'savings_account',
    'ind_cco_fin_ult1': 'current_account',
    'ind_cder_fin_ult1': 'derivada_account',
    'ind_cno_fin_ult1': 'payroll_account',
    'ind_ctju_fin_ult1': 'junior_account',
    'ind_ctma_fin_ult1': 'mas_particular_account',
    'ind_ctop_fin_ult1': 'particular_account',
    'ind_ctpp_fin_ult1': 'particular_plus_account',
    'ind_deco_fin_ult1': 'short_term_deposit',
    'ind_deme_fin_ult1': 'medium_term_deposit',
    'ind_dela_fin_ult1': 'long_term_deposit',
    'ind_ecue_fin_ult1': 'e-account',
    'ind_fond_fin_ult1': 'funds',
    'ind_hip_fin_ult1': 'mortgage',
    'ind_plan_fin_ult1': 'pensions',
    'ind_pres_fin_ult1': 'loans',
    'ind_reca_fin_ult1': 'tax_payments',
    'ind_tjcr_fin_ult1': 'credit_card',
    'ind_valo_fin_ult1': 'securities',
    'ind_viv_fin_ult1': 'home_account',
    'ind_nomina_ult1': 'payroll',
    'ind_nom_pens_ult1': 'pension',
    'ind_recibo_ult1': 'direct_debit'
}

# Applying the column name changes in data
train_data.rename(columns=column_mapping, inplace=True)
```

Data Types

All the was object. some columns data types changed from object to float.

```
ı [16]:  print(train_data.dtypes)

record_date                      object
customer_id                      object
employee_status                  object
country_of_residence             object
gender                           object
age                             float64
customer_since                   object
new_customer_index               object
seniority_months                float64
primary_relationship_type        object
last_primary_relationship        object
customer_type_last_month         object
residence_flag                   object
foreigner_flag                   object
customer_acquisition_channel     object
deceased_flag                    object
address_type                     object
province_code                    object
province_name                    object
active_customer_flag             object
household_income                float64
customer_segment                 object
savings_account                  object
```

```
savings_account                  object
current_account                  object
derivada_account                 object
payroll_account                  object
junior_account                   object
mas_particular_account           object
particular_account               object
particular_plus_account          object
short_term_deposit               object
medium_term_deposit              object
long_term_deposit                object
e-account                        object
funds                            object
mortgage                         object
pensions                         object
loans                            object
tax_payments                     object
credit_card                      object
securities                       object
home_account                     object
payroll                         float64
pension                         float64
direct_debit                      int64
dtype: object
```

Data Cleaning

```
In [21]:  #deleting unnesesary column
          train_data.dropna(subset=["customer_acquisition_channel"], inplace=True)
```

Missing Value

```
n [25]:  train_data['last_primary_relationship'].fillna('Unknown', inplace=True)
         train_data['customer_type_last_month'].fillna('Unknown', inplace=True)
```

```
n [26]:  import warnings
         warnings.simplefilter("ignore")
```

```
n [27]:  train_data['payroll'].fillna(train_data['payroll'].mode()[0], inplace=True)
         train_data['pension'].fillna(train_data['pension'].mode()[0], inplace=True)
```

```
n [28]:  import warnings
         warnings.simplefilter("ignore")
```

```
n [29]:  train_data['province_name'] = train_data['province_name'].fillna(method='ffill')
```

Outlier Detection